Quantitative Analysis of Volatile Organic Compounds Using Ion Mobility Spectra and Cascade Correlation Neural Networks

Peter de B. Harrington* **and Peng Zheng**
Ohio University
Center for Intelligent Chemical Instrumentation
Department of Chemistry
Clippinger Laboratories
Athens, OH 45701-2979

Dennis M. Davis
U.S. Army Edgewood Research, Development and Engineering Center
Chemical and Biological Detection Research Team
Aberdeen Proving Ground, MD 21010-5423

ABSTRACT

Ion mobility spectrometry (IMS) is a powerful technique for trace organic analysis in the gas phase. Quantitative measurements are difficult, because IMS has a limited linear range. Factors that may affect the instrument response are pressure, temperature and humidity. Nonlinear calibration methods, such as neural networks, may be ideally suited for IMS. Neural networks have the capability of modeling complex systems. Many neural networks suffer from long training times and overfitting. Cascade correlation neural networks train at very fast rates. They also build their own topology, that is number of layers and number of units in each layer. By controlling the decay parameter in training neural networks, reproducible and general models may be obtained.

INTRODUCTION

IMS is a powerful technique for analysis of trace quantities of organic compounds in the gas phase. IMS is inexpensive, offers high sensitivity, fast response and low detection limits. Sophisticated computer software coupled to an IMS instrument may allow a synergism to be obtained that will generate an intelligent or smart instrument. These instruments will incorporate expertise in data interpretation and processing as part of the hardware. There is great potential for automated identification and quantitation of volatile substances by IMS. The suitability of neural networks for IMS data analysis has already been demonstrated[1,2]. Automated systems may be developed for identification and classification of unknown materials.

A general system was presented for semi-quantitative measurements of volatile organic compounds in air that used backpropagation neural networks (BNNs)[3]. Although these systems are popular tools, they were abandoned for this project because of several disadvantages. The most important limitation is that BNNs train at slow rates. For difficult cases, BNNs may not converge to a satisfactory error. In addition, determining the network

*Corresponding author.

configuration (i.e., the number of layers and processing units) is a difficult and important step for building reliable models. The cascade correlation network (CCN) learns its own network configuration during training.[4] CCNs train faster than BNNs by adjusting a single unit at a time. The result is a network that configures its topology and trains at an astonishing rate. CCNs are valuable tools for obtaining general calibration models from IMS data. A CCN that identifies and quantitates a variety of organic solvents over a broad range of concentrations has been developed.

An IMS instrument consists of a sample inlet system, a reaction region, a drift region and an ion collector. The carrier gas (air or nitrogen) sweeps the sample into the reaction region where the sample gas is ionized. The ions formed in the reaction region are injected into the drift region where the time for ions to move through an applied electric field is measured. The drift velocity, V_d, of an ion is obtained by dividing the length of the drift region, L, by the drift time, t_d. The ion mobility, K, is obtained by dividing the drift velocity by the accelerating potential, E. The relationships are given in equations (1) and (2).

$$V_d = \frac{L}{t_d} \tag{1}$$

$$K = \frac{V_d}{E} \tag{2}$$

The mobility value may be corrected to standard gas density, 2.687 x 10^{19} molecule/cm that corresponds to standard conditions of 273 Kelvin and 760 torr. The reduced mobility K_0 is obtained by

$$K_0 = K\left(\frac{P}{760}\right)\left(\frac{273}{T}\right) \tag{3}$$

for which P and T are the experimental pressure and temperature, respectively.

Factors that affect K_0 include imprecise determination of the electric field, imprecise temperature measurements of the drift space, temperature gradients, pressure fluctuation, and humidity of the drift gas. Pressure fluctuations are important when the drift gas flow rate is pump controlled. Variable water content of the drift gas affects the water content of the clustered reactant ions. Sample introduction may saturate the system and contaminants or multiple analytes may interact in a complex manner. Under saturated conditions, multiple peaks due to contaminants in the sample, oligomer formation, and ion-molecule reactions in the drift space are common. Also K_0 values are dependent on the drift gas. In general, K_0 changes as a function of the polarizability, α, of the drift gas.

The drift tube of the instrument is continually swept with air or nitrogen, to clear neutral sample molecules from the drift tube, and to serve as the drift gas of the instrument.

The most common form of primary ionization is by β-particle radiation from a ^{63}Ni foil in the reaction region of the IMS cell. A series of cascading reactions progresses and results in the formation of a set of reactant ions. Sample vapors are introduced into the reaction region via a carrier gas where reactant ions collide with sample vapor molecules and product ions are formed. The predominant mechanism of formation for positive ions is proton transfer from $(H_2O)_nH^+$ reactant ions to the sample molecule, which is followed by a number of possible complex reactions. Typical reactions are with additional sample molecules present, reactions with water vapor molecules, and reaction with other neutral molecules present in the reaction region of the mobility spectrometer. Product ions containing one analyte molecule are referred to as "monomers", two analyte molecules, "dimers", and so forth.

Ions are injected into the drift region of the instrument using an electric shutter. Characteristic constant velocities are attained for particular ions which form a near-Gaussian distribution during drift tube traversal. At the end of the drift region, ions strike a collector electrode (Faraday cup) and generate a weak source signal (typically 10-100 pA) at a frequency of 30-40 Hz.

EXPERIMENTAL SECTION

The ion mobility spectrometer used to obtain spectra was a hand-held IMS device, the Airborne Vapor Monitor, AVM, (Graseby Ionics, Ltd. Watford, Herts, U.K.). Spectra from the ion mobility spectrometer were collected on an IBM-compatible 80386 25MHz personal computer using Advanced Signal Processor Software (ASP, Graseby Ionics, Ltd.). Typical data collection parameters are given in Table I.

Table I. Typical Operating Conditions for the AVM

Ionization Source	^{63}Ni
Gating Pulse Repetition Rate	33 Hz
Cell Temperature	30 °C
Cell Pressure	760 torr
Drift Gas	Clean dry air
Drift Gas Flow	400 mL/min
Spectral Mode	Positive Ion
Points Per Spectrum	640
Sampling Frequency	33 KHz
Delay to Start of Sampling	0 μs

Data acquisition is initiated at the midpoint of the gating pulse. A spectrum is collected at a user specified sampling frequency. In practice, an analog signal is converted with a flash analog-to-digital converter (ADC) which has twelve bit resolution. The spectra

are stored in binary format and can be converted to ASCII format as required.

All laboratory vapors were generated using a vapor generator called the Q5. The Q5 generator, Figure 1, has 15 component parts. These parts are: (1) an equilibrator assembly, (2) an air supply stopcock, (3) a constant pressure regulator (stabilizer) for the air supply, (4) a flowmeter (manometer) for the air supply, (5) a constant pressure regulator (stabilizer) for the diluent air supply, (6) stopcocks for the stabilizers, (7) a stopcock for the flow of air from the equilibrator to the mixing chamber, (8) a flowmeter (manometer) for the diluent air supply, (9) a mixing chamber, (10) a reservoir, (11) a reservoir exhaust stopcock, (12) sampling stopcocks, (13) a charcoal trap on the exhaust of the reservoir, (14) a charcoal canister on the sampling line, (15) and sampling bubblers.

The equilibrator assembly is the liquid test reagent container of the dilution apparatus. Dry air, under a constant controlled pressure, flows into the equilibrator. This air stream passes over the surface of the test reagent, and becomes saturated with the reagent vapor. The dry air-test vapor mixture flows from the equilibrator assembly to the mixing chamber where it is diluted with dry air to the required concentration of milligrams test vapor per liter of dry air. The equilibrator is maintained at a constant temperature by partial immersion in a constant temperature water bath. The water bath is maintained at 25 °C. Included in the equilibrator is a porous alundum oxide cylinder to produce a greater surface area for the liquid-vapor equilibration.

The flow of air through the equilibrator is controlled by a stopcock in the air supply line, a constant pressure regulator for the air supply, and a flowmeter. The stopcock is located at the inlet of the equilibrator, and acts as the shutoff valve for the air supply, from the flowmeter to the equilibrator. The constant pressure for the air supply is maintained by bubbling the dry air through a constant level of fluid, e.g., water, in the stabilizer. By raising or lowering the level of the fluid in the stabilizer, the pressure of the air supply is increased or decreased. The level of the fluid is raised by adding fluid to the stabilizer, and lowered by draining fluid through the stabilizer stopcock that is located on the bottom of the stabilizer. Changing the pressure of the air supply in this way increases or decreases the flow of the test vapor through the dilution apparatus. Excess air that passes through the stabilizer is vented to the laboratory hood. The flowmeter measures the flow rate of the dry air-test vapor mixture in milliliters per minute. The flow rate of the diluent air is controlled in the same way as the equilibrator air supply. The flowmeter for the diluent air is measured in liters per minute. The nominal concentration of the test vapor can be calculated by

$$C = \frac{fp}{(F+f)P} \tag{4}$$

for which C is the concentration of the test vapor in parts per million, f is flow rate of air through the equilibrator, F is the flow rate of the diluent air, p is the vapor pressure of the test reagent at the temperature of the experiment, and P is the atmospheric pressure. Thus, the concentration of the test vapor may be easily changed by varying either the flow rate of air through the equilibrator or the flow rate of the diluent air. In practice, it works best to change the flow rate of the diluent air, when possible, because the efficiency of the vapor generation in the equilibrator decreases as the flow rate increases.

The dry air-test vapor mixture from the equilibrator and the diluent air are passed into the mixing chamber located at the entrance of the reservoir. The dilute test vapor is thoroughly mixed by a swirling circular motion of the air in the mixing chamber before entering the reservoir. The reservoir is the container for the diluted test vapor, from which samples are taken for concentration analysis and for testing purposes. There is a charcoal canister located on the exhaust of the reservoir. This canister serves as a scrubber to remove test vapors passing from the reservoir to the atmosphere in the laboratory hood. There are two sampling ports on the reservoir, one for taking samples for concentration analysis; the other for removing the vapor for testing purposes.

IMS data was collected for the following solvents and concentration ranges given in Table II. This set of data was split almost equally into calibration and prediction sets of data with the following exception. The prediction set had 9 more reactant ion spectra for which solvent vapors were present during data collection.

Table II. Spectra used in the Evaluation (Number of Calibration:Prediction).

Acetic Acid	100 ppb (3:3)	1 ppm (4:4)	10 ppm (5:4)		
Acetic Anhydride	100 ppb (2:3)	1 ppm (5:4)	5 ppm (4:4)		
Acetone	1 ppm (3:3)	10 ppm (2:2)	100 ppm (3:4)	1000 ppm (10:8)	
Diethyl Ether	4 ppm (4:4)	40 ppm (3:3)	400 ppm (4:4)		
Isooctane	3.5 ppm (6:6)	35 ppm (6:5)	350 ppm (3:3)		
Isopropanol	4 ppm (4:4)	40 ppm (2:2)	400 ppm (2:2)		
Methyl Benzoate	5 ppb (3:2)	10 ppb (5:5)	20 ppb (5:5)	50 ppb (3:2)	100 ppb (3:3)
o-Nitro Toluene	2.5 ppb (5:4)	25 ppb (9:10)	250 ppb (4:3)		
Toluene	15 ppb (3:2)	230 ppb (3:3)	1.7 ppm (3:4)	10 ppm (4:5)	
Reactant Ions	(29:38)				

All data analyses were run on a 60 MHz Pentium PC operating under DOS 6.2. Code was developed with Watcom's 32 bit C++ compiler version 9.5b. The cascade correlation program is shareware that is available from Carnegie Mellon University. The program was modified to use an improved random number generator. The ANSI C 16 bit random number

generator should be avoided for neural network programs.[5]

The IMS data were corrected for baseline by subtracting the average intensity from the range of 1.5-3.0 ms. The spectra were reduced by removing all the points less than 3.0 ms and greater than 16.0 ms. For CCN evaluations, each spectrum was scaled with respect to intensity by dividing every point in the spectrum by 1000. After preprocessing each spectrum was composed of 390 single precision floating point intensities.

The CCNs were trained with a sigmoid prime offset of 0.1. The output learning rate was 0.1 and the hidden unit learning rate was 0.001. For two CCN training configurations, the decay rates were either a factor of ten greater than the learning rates, or zero. The output units were trained for a maximum of 400 epochs and the hidden units were trained for a maximum of 200 epochs. If the error decreased less than 0.1% or 3% in 12 epochs, then training terminated for the hidden and output units, respectively. Twenty hidden units were trained simultaneously with the one yielding the highest correlation being used. Both output and hidden units were trained with a quickprop maximum step threshold value (μ) of 2.0. The networks were trained until the relative training error (i.e., error divided by the variance of the outputs) was 5%.

DATA ANALYSIS

Nine dependent variables were obtained for each spectrum. Each variable corresponds to the identity of a compound in the training set. The concentration values were transformed to p-values (i.e., $-\log_{10}$). Reactant ions were included from each compound acquisition. Their p-values were arbitrarily assigned values of 10, which corresponds to an arbitrary detection limit of the instrument. This transformation is necessary so that the concentration values will be equally weighted. Otherwise the networks must recognize small differences in concentration that result in bivariate predictions of either high or low concentration. This problem occurs because the differences between low concentration design points will be less than the variance about the high concentration design point. All the dependent variables are assigned values of 10, except that the variable corresponding to the compound is set to the p-concentration value. This approach is beneficial, because the identity of a compound is given by the prediction variable with the smallest output. The outputs correspond to concentrations, and spectra of multiple compounds may be modeled without modification.

Neural networks are powerful pattern recognizers and are susceptible to overfitting training data. It is always a good idea to implement control methods of pattern recognition for evaluation of a neural network's performance. Partial least squares (PLS) was chosen as a control methods.[6] The PLS model was constructed so that the number of latent variables furnished a minimum error for prediction set of data.

Errors were reported as standard error prediction (SEP) and relative standard error of prediction (RSEP) as given below.

$$SEP=\sqrt{\frac{\sum_{j=1}^{p}\sum_{i=1}^{n}(Y_{i,j}-\hat{Y}_{i,j})^2}{np}} \tag{5}$$

$$RSEP=\sqrt{\frac{\sum_{j=1}^{p}\sum_{i=1}^{n}\left(\frac{Y_{i,j}-\hat{Y}_{i,j}}{Y_{i,j}}\right)^2}{np}} \tag{6}$$

The SEP has the same units as the prediction variables which are p-concentrations and RSEP is unitless and is reported as a percentage. The number of outputs in the network is given by p and the number of observations in the prediction set is given by n. The prediction set is an external set of validation data. The predicted value is $\hat{y}_{i,j}$ and the target output is $y_{i,j}$ for the j output and i observation.

The CCN is an alternative method to the more popular BNN. The differences between the two methods are that a CCN trains at a substantially faster rate and CCNs determine their own topology (i.e., number of layers and units). CCNs train by adjusting only a single unit at a time. Because CCNs fix all previously trained units, they have the capability of incremental learning. Therefore, if a new compound needs to be added to the calibration, the network can be added to its existing structure and save time during training.

CCNs like other neural networks train by adjusting weights so that the training error is minimized. Like any nonlinear optimization, the procedure is iterative and dependent on the initial conditions. For neural networks, the initial conditions are the random weight values assigned to weights prior to training. It has been shown that these initial conditions introduce a modeling error into neural network models.[7] Therefore all neural network results should be accompanied by figures of merit indicating this lack of precision.

The CCNs initiate training with only output units. These units are adjusted until a minimum prediction error is obtained. A candidate unit is added as hidden unit. The unit is trained so that the absolute value of the covariance between the unit's output and the residual error is maximized. The covariance is given as

$$s_j=\sum_{k=1}^{p}\left|\sum_{i=1}^{n}(o_{i,j}-\bar{o}_j)(e_{i,k}-\bar{e}_k)\right| \tag{7}$$

for which s_j is the covariance of unit j. The number of output units is given by p, and n is the number of observations in the training set. The average output per epoch is given by $\bar{o}_j$ and the average error per epoch is given by $\bar{e}_k$. After a hidden unit's covariance is maximized, its weights are fixed. The output of the hidden unit is connected to the output units, and the output units are adjusted until the prediction error is minimized. The next hidden unit is added so that the previous hidden unit feeds into it. Therefore, the hidden units have a

cascading architecture because each new unit receives the outputs from all the previously added hidden units as well as the inputs, i.e., the spectra. The cascading architecture allows the CCN to model higher order functions that would be impossible for backpropagation networks. Because only one unit is adjusted at a time, the other units in the network can cache their outputs, and the network can train with fewer computations.

The networks presented in this study are all trained in batch mode with decay and sigmoid prime offsets. The basic training model for adjusting component *j* of weight vector for output unit *k* is given by

$$\Delta w_{j,k}(t+1) = -dw_{j,k}(t) + \eta \sum_{i=1}^{n} (y_{i,k} - \hat{y}_{i,k}) x_{i,j} \qquad \textbf{(8)}$$

for which *t* indicates training cycle (i.e., one pass through the training set). A training cycle is a complete pass through the *n* spectra in the training set, and is referred to as an epoch. A decay parameter (*d*) causes weights that are not pertinent (i.e., independent of prediction error) to decrease in magnitude. The learning rate is designated as η, $y_{i,k}$ is the target output and $\hat{y}_{i,k}$ is the predicted output of unit *k* and observation *i*, $x_{i,j}$ is the *j* component of input *i*. The training errors are summed for the number of observations in the training set (*n*). The hidden units are adjusted in a similar process.

$$\Delta w_{j,k}(t+1) = -dw_{j,k}(t) + \eta \sum_{i=1}^{n} s_i x_{i,j} (f'(net_{i,k}) + spo) \qquad \textbf{(9)}$$

The hidden units have sigmoid activation functions. Therefore the first derivative of the sigmoid function, $f'(net_{i,k})$, is needed for the weight adjustment. Network paralysis may be minimized by the sigmoid prime offset parameter (*spo*). This parameter propagates errors back through units whose sigmoid derivatives are zero or small. The value of s_i is given by

$$s_i = \sum_{k=1}^{p} \sigma_k [(y_{i,k} - \hat{y}_{i,k}) - \overline{(y_{i,k} - \hat{y}_{i,k})}] \qquad \textbf{(10)}$$

for which σ_k is the sign of the covariance of a hidden unit and the residual error of an output unit *k* for a training epoch. The other terms are the residual error and the average residual error for an entire training epoch. The CCN network uses the quickprop algorithm which is beyond the scope of this paper[8].

An important parameter for neural network generalization is the decay parameter. If this parameter is large, spurious weight values may be removed and the networks avoid overfitting the data. The networks may model large variances in the training data. However, if this value is too large the networks may never converge, because the weights will decay to zero. Increasing the decay parameter also increases training time. The decay and training time for the CCN evaluations are given in Table III.

Discussion of Results

The inclusion of data below the detection limit of an instrument is often avoided in developing calibration models. If data below the detection limit is included, nonlinearities in the model may be introduced. Neural networks can accomplish nonlinear modeling. For example, changes in concentration below the detection limit may result in no change in signal. However, recognition of the absence of a sample is important. Therefore, background scans were used for both calibration and prediction. An arbitrary detection limit was set to 0.1 ppb, and this value was assigned to the background scans. Therefore, recognition of a sample is achieved by predictions greater than this minimum concentration.

PLS is used as the control method. From the monotonic decrease of calibration error in Figure 2, one can see that PLS models nonlinear relationships. However, after 35 latent variables are used in the model, the error in predicting an external set of data increases. A PLS model with 35 latent variables is used as a control for comparison with CCN predictions. The PLS predictions are biased in that the model was optimized with respect to the prediction set. The optimal SEP and RSEP from this model are 0.322 (p-conc.), and 4.1%, respectively.

The effect of the decay parameter was evaluated with two CCN training configurations that were identical except for the use of a decay parameter. One configuration did not use weight decay in training, and the other had the weight decay set to an order of magnitude greater than learning rate. The magnitude of the decay parameter was chosen to be as large as possible, yet still furnish a reduction in error for each training cycle. Each training configuration was used to construct five networks which differed in only the random initial weights. Five replicates are used to study modeling error, caused by different initial conditions for the nonlinear optimization of the network.

Table III. Summary of CCN Results

Evaluation	SEP d/η=10 (p-Conc.)	RSEP d/η=10 (%)	Number of Hidden Units	SEP d/η=0 (p-Conc.)	RSEP d/η=0 (%)	Number of Hidden Units
Run 1	0.13	1.8	114	1.1	13	6
Run 2	0.14	1.9	119	1.05	13	6
Run 3	0.13	1.8	125	1.04	12	6
Run 4	0.15	2.0	119	1.06	13	6
Run 5	0.14	1.9	126	1.06	13	5
Average	**0.14±0.01**	**1.9±0.1**	**121±5**	**1.06±0.02**	**12.8±0.5**	**5.8±0.5**
PLS	0.322	4.1				

Large decay values increase network training times. The networks without the weight decay term took approximately 40 minutes each to train on a personal computer. The networks that trained with the large weight decay values each took approximately 15 hours to train. The increase in time by using large weight decay values, also applies to

backpropagation networks, which would be on the order of several days. Note that the number of hidden units is approximately 20 times larger for the CCN with the high decay value.

Large decay values are advantageous in that the networks avoid overfitting. A prevailing belief among neural network practitioners is that the neural network should be as small as possible to minimize overfitting the data set with too many adjustable parameters. These results demonstrate that including more adjustable parameters (i.e., hidden units) may improve a network's ability to generalize as shown by a decrease in error with external prediction sets. Note that the errors with the network that did not use a decay parameter are six times larger, and the lack of precision caused by modeling error is five times greater. The external prediction set was not used in training the network, or selecting training parameters.

Figures 3A-3I give the predicted and measured concentrations for each compound class. The predicted concentrations are the outputs from each of the nine output units. The points in the figures correspond to the spectra in the prediction set. This data was obtained from the results of five replicate runs of the CCN with the decay parameter. The CCN data is given as 95% confidence levels for each point, which indicates the model error associated with a spectrum. Spectra with large modeling errors are indicative of outliers. The PLS predictions are included in the figures. The reference line in the figures gives the ideal response. The points at the reported p-concentration of 10 either are reactant ions or are different compounds from those of the compound class. Note that the CCN performs favorably when compared to PLS. Each output unit corresponds to a compound in the training set. A compound is identified by its smallest p-concentration value. No compounds were misidentified in these evaluations.

Conclusions

IMS is an exciting area for the application of neural networks. The CCN is an excellent tool for the chemist. It has the advantages of a self-configuring topology, fast training rate and incremental learning. The use of the decay parameter can help a neural network generalize by modeling larger variations in the spectra. Large decay values not only reduce modeling errors, but also reduce the network convergence rate. If training time is increased by the use of the decay parameter, networks that train efficiently are desirable.

This study presents a general model for both quantitation and identification of organic compounds by IMS. Neural networks can perform better than standard data analysis methods such as PLS. Neural networks are well suited for IMS because they can be trained to have a property of shift invariance. This property is important to IMS data analysis for which temperature, pressure and other atmospheric fluctuations may cause peak shifting in the spectra. Incremental learning with CCNs and the study of mixtures by IMS are areas of future research.

Acknowledgements

The U.S. Army E.R.D.E.C. is thanked for collecting and furnishing the data. The conference organizers and participants are thanked for the helpful discussions and comments. Lijuan Hu and Busolo Wa Wabuyele are thanked for their help with writing this paper.

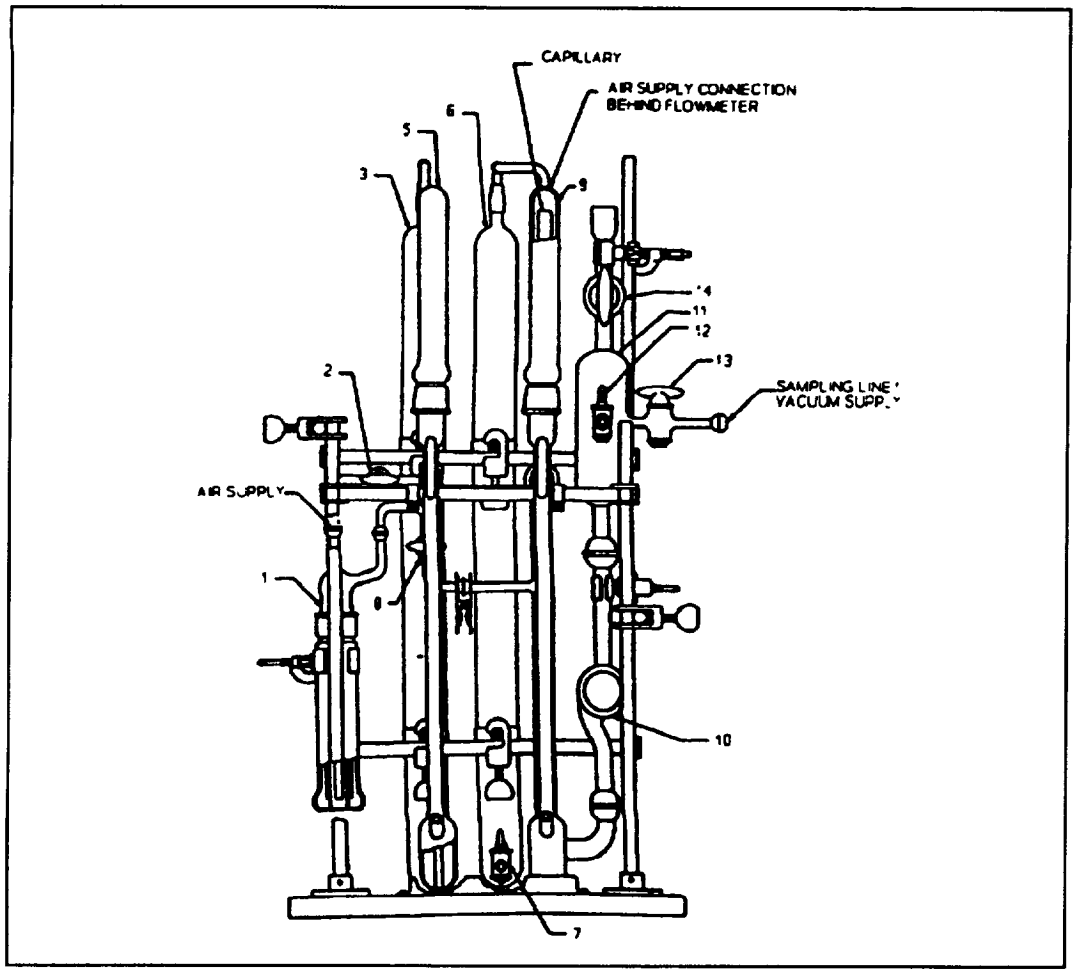

Figure 1. Schematic of the Q5 vapor generator.

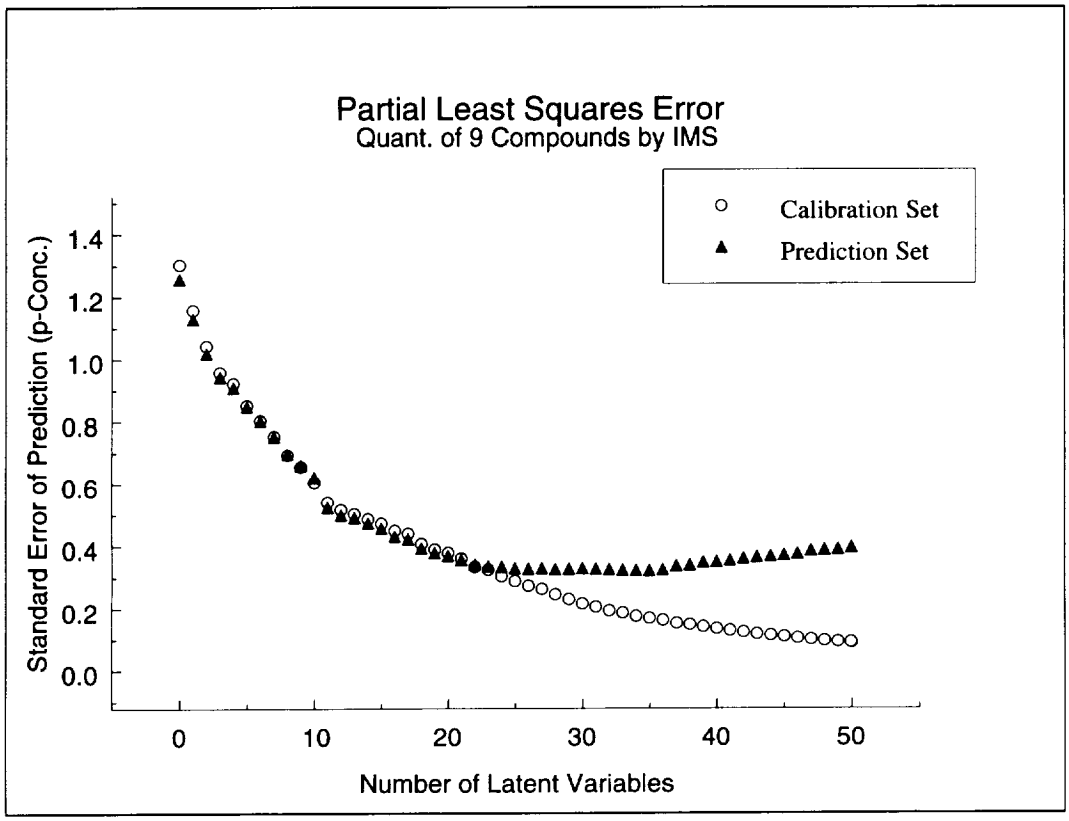
Partial Least Squares Error
Quant. of 9 Compounds by IMS
Calibration Set
Prediction Set
Standard Error of Prediction (p-Conc.)
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
Number of Latent Variables

Figure 2.

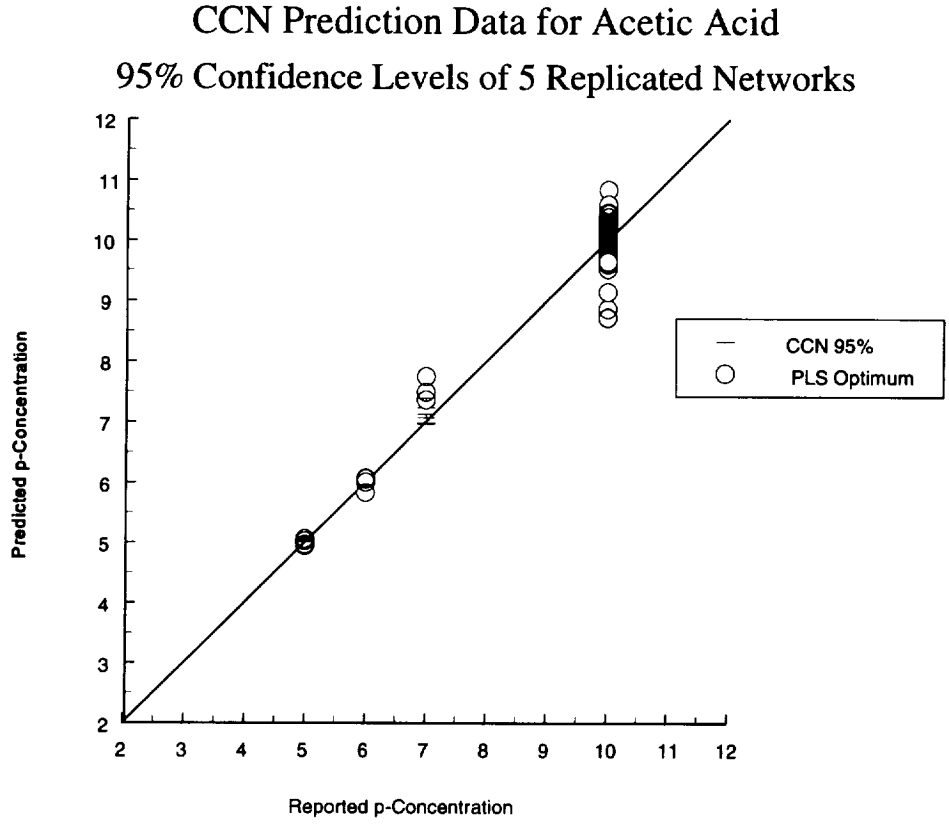
CCN Prediction Data for Acetic Acid
95% Confidence Levels of 5 Replicated Networks
Predicted p-Concentration
Reported p-Concentration
CCN 95%
PLS Optimum
2
3
4
5
6
7
8
9
10
11
12

Figure 3A.

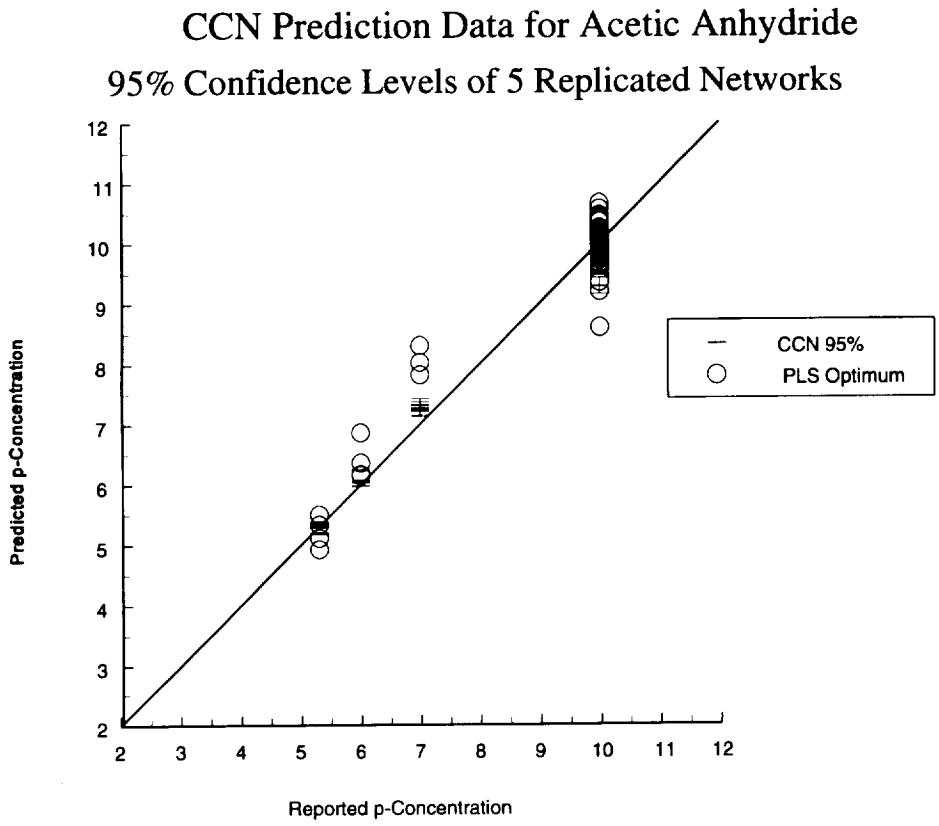
CCN Prediction Data for Acetic Anhydride
95% Confidence Levels of 5 Replicated Networks
Predicted p-Concentration
Reported p-Concentration
CCN 95%
PLS Optimum
2
3
4
5
6
7
8
9
10
11
12

Figure 3B.

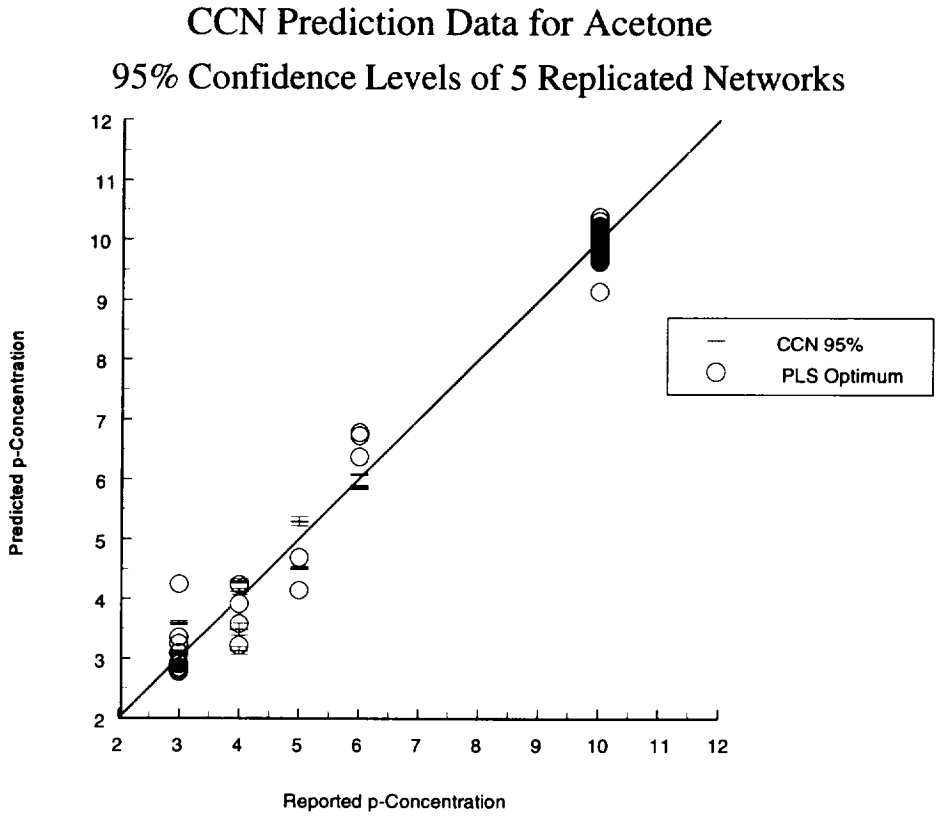
CCN Prediction Data for Acetone
95% Confidence Levels of 5 Replicated Networks
Predicted p-Concentration
Reported p-Concentration
CCN 95%
PLS Optimum
2
3
4
5
6
7
8
9
10
11
12

Figure 3C.

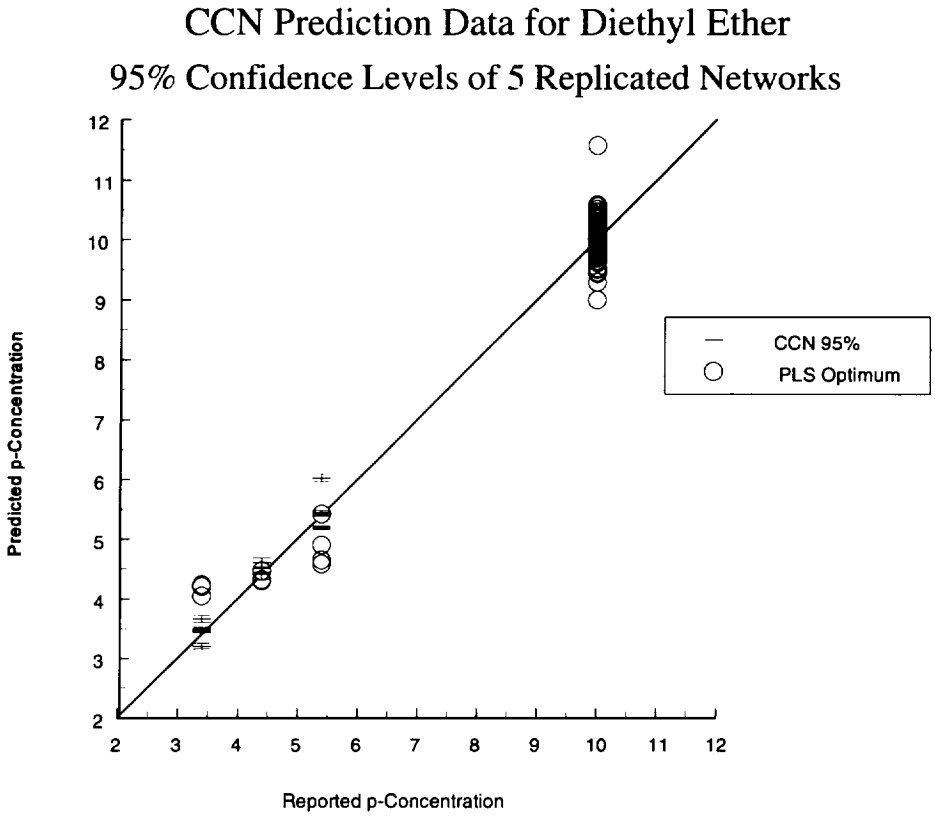
CCN Prediction Data for Diethyl Ether
95% Confidence Levels of 5 Replicated Networks
Predicted p-Concentration
Reported p-Concentration
CCN 95%
PLS Optimum
2
3
4
5
6
7
8
9
10
11
12

Figure 3D.

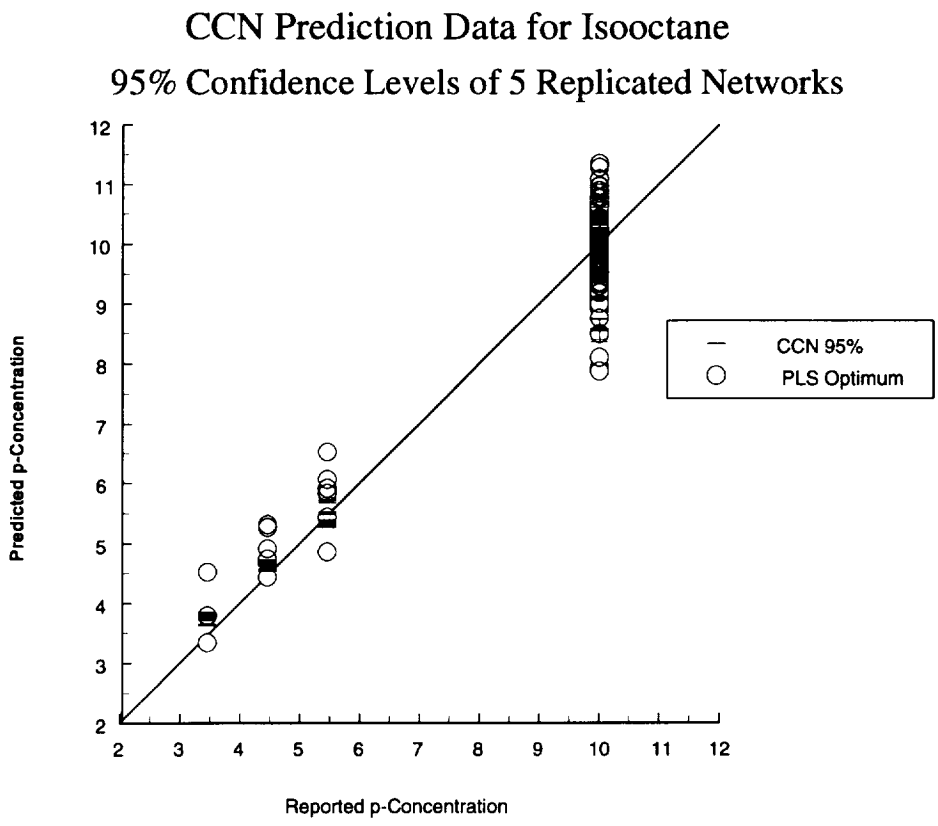
CCN Prediction Data for Isooctane
95% Confidence Levels of 5 Replicated Networks
Predicted p-Concentration
Reported p-Concentration
CCN 95%
PLS Optimum
2
3
4
5
6
7
8
9
10
11
12

Figure 3E.

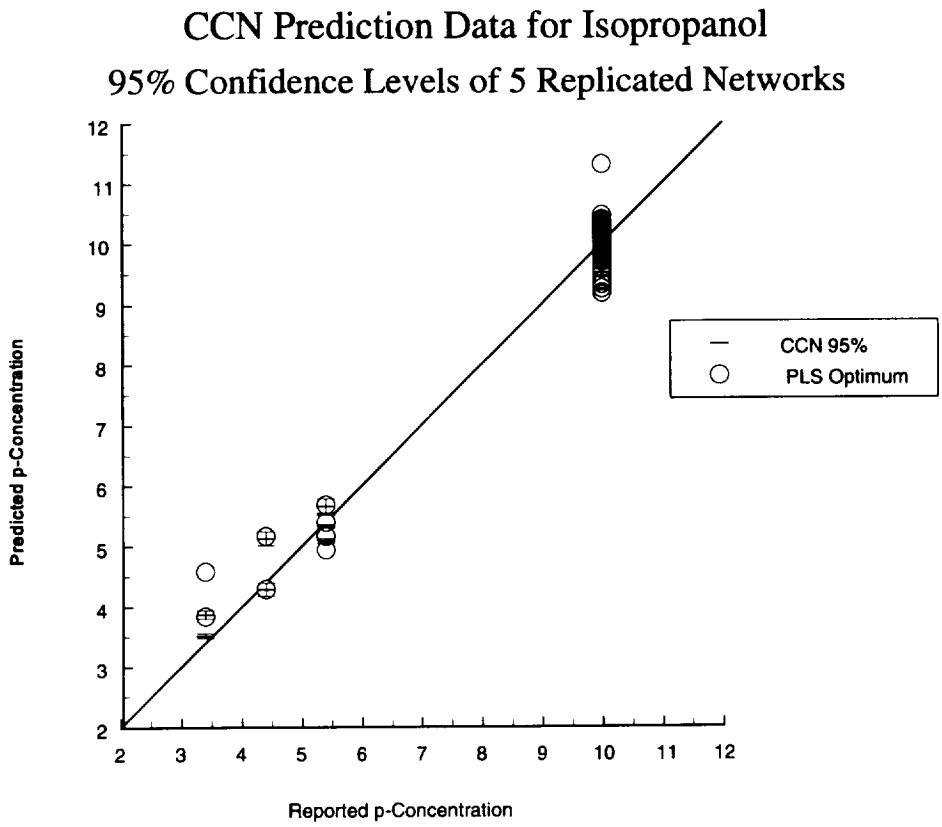
CCN Prediction Data for Isopropanol
95% Confidence Levels of 5 Replicated Networks
Predicted p-Concentration
Reported p-Concentration
CCN 95%
PLS Optimum
2
3
4
5
6
7
8
9
10
11
12

Figure 3F.

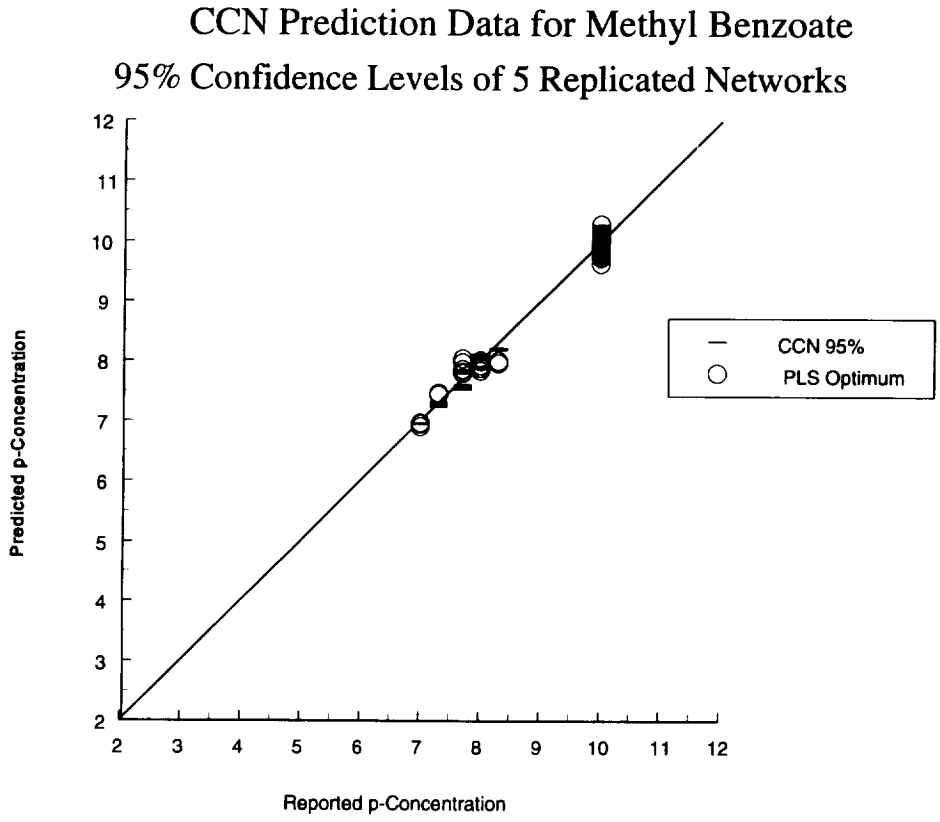
CCN Prediction Data for Methyl Benzoate
95% Confidence Levels of 5 Replicated Networks
Predicted p-Concentration
Reported p-Concentration
CCN 95%
PLS Optimum
2
3
4
5
6
7
8
9
10
11
12

Figure 3G.

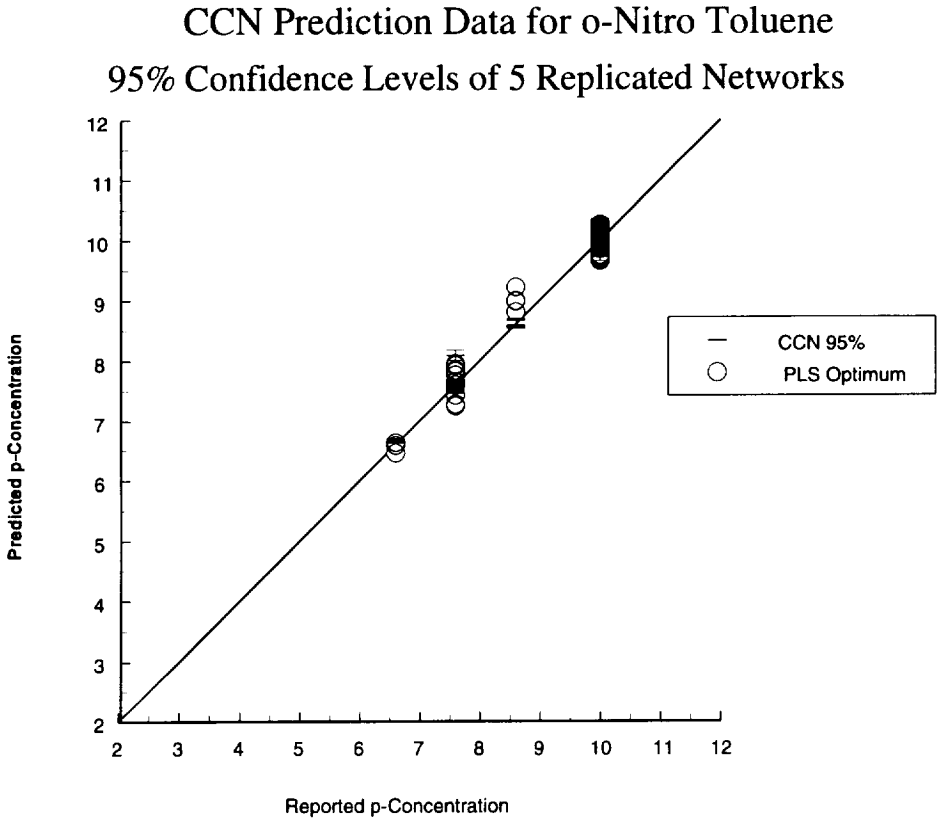
CCN Prediction Data for o-Nitro Toluene
95% Confidence Levels of 5 Replicated Networks
Predicted p-Concentration
Reported p-Concentration
CCN 95%
PLS Optimum
2
3
4
5
6
7
8
9
10
11
12

Figure 3H.

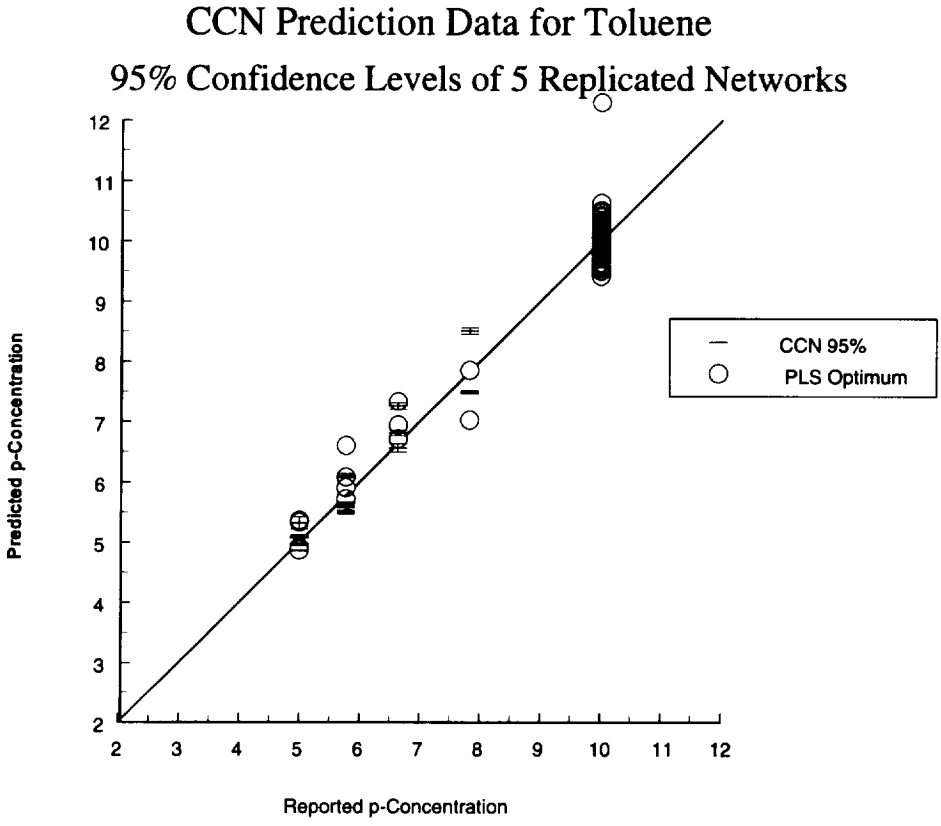
CCN Prediction Data for Toluene
95% Confidence Levels of 5 Replicated Networks
Predicted p-Concentration
Reported p-Concentration
CCN 95%
PLS Optimum
2
3
4
5
6
7
8
9
10
11
12

Figure 3I.

Citations

1. Bell, S.E., Mead, W.C., Jones, R.D., Eiceman, G.A., and Ewing, R.G. Connectionist Hyperprism Neural Network for the Analysis of Ion Mobility Spectra: An Empirical Evaluation. *J. Chem. Inf. Comput. Sci.*, **33**, 609-615 (1993).

2. Boger, Z. and Karpas, Z. Use of Neural Networks for Quantitative Measurements in Ion Mobility Spectrometry (IMS), *J. Chem. Inf. Comput. Sci.*, **34**, 576-580 (1994).

3. Zheng, P. Davis, D.M. and Harrington, P.B. Comparison of Backpropagation and Counterpropagation Neural Networks for Quantitative Analysis of Ion Mobility Spectra presented at the 1994 Pittsburgh Conference, Chicago, IL, March 1994, **641**.

4. Fahlman, S.E. and Lebiere C., The Cascade Correlation Architecture. Carnegie Mellon University Report **CMU-CS-90-100**, August 1991, 1-13.

5. Masters T. Practical Neural Network Recipes in C++. Academic Press, San Diego, CA, pp. 128-132.

6. Geladi, P and Kowalski, B.R. Partial Least-Squares Regression: A Tutorial. *Anal. Chim. Acta,* **185**, 1-17 (1986).

6. Harrington, P.B. Temperature-Constrained Backpropagation Neural Networks. *Anal. Chem.*, **66**, 802-807 (1994).

8. Fahlman, S.E. An Empirical Study of Learning Speed in Back-Propagation Networks. Carnegie Mellon University Report **CMU-CS-88-162**, September 1988, 1-17.